Управление наукой и наукометрия. 2022. Т. 17, № 3.
Science Governance and Scientometrics. 2022;17(3).
ISSN 2686-6706 (print), ISSN 2686-6714 (online)

ИНФОРМАЦИОННАЯ СРЕДА НАУЧНОЙ, ОБРАЗОВАТЕЛЬНОЙ И ИННОВАЦИОННОЙ СФЕР

INFORMATION ENVIRONMENT OF SCIENCE, EDUCATION AND INNOVATION

Научная статья / Original research
УДК 002.63
https://doi.org/10.33873/2686-6706.2022.17-3.380-407

Национальная научно-информационная инфраструктура: проблемы, задачи и перспективы

Андрей Евгеньевич Гуськов[1]✉, Александр Сергеевич Карауш[2], Илья Евгеньевич Меньщиков[3], Андрей Вячеславович Школин[3], Виталий Олегович Недельский[4], Денис Шамилевич Сабиров[5], Тимур Николаевич Щукин[6]

[1] Российский научно-исследовательский институт экономики, политики и права в научно-технической сфере (РИЭПП), г. Москва, Россия
[2] Государственная публичная научно-техническая библиотека России, г. Москва, Россия
[3] Институт физической химии и электрохимии им. А. Н. Фрумкина Российской академии наук, г. Москва, Россия
[4] Горно-Алтайский государственный университет, г. Горно-Алтайск, Россия
[5] Институт нефтехимии и катализа Уфимского федерального исследовательского центра РАН (ИНК УФИЦ РАН), г. Уфа, Россия
[6] Московская школа управления «Сколково», Московская обл., Россия
✉ *guskov.andrey@gmail.com*

Резюме

Введение. Большие объемы научных данных и развитие парадигмы Data-Driven Science формируют потребность в принципиально новых инструментах для работы с информацией и научных коммуникаций. Появление таких инструментов возможно только при наличии технологического фундамента — развитой цифровой инфраструктуры для хранения и обмена большого разнообразия научных данных, ориентированной на использование не только исследователями, но и программными агентами. В статье анализируются проблемы сложившихся в России практик управления научной информацией и существующих барьеров, которые препятствуют созданию инструментов нового поколения. Их преодоление является очень сложной и масштабной задачей, что обусловлено разнообразием акторов научно-информационной деятельности и видов научной информации, спецификой различных предметных областей и другими факторами. Для ее решения необходима Стратегия развития Национальной научно-информационной инфраструктуры — комплекс системных изменений стандартов и практик

работы с научной информацией, затрагивающих значительную часть научного сообщества. **Результаты и дискуссия.** Сформулированы основные положения Стратегии, включая цель, задачи и принципы разработки, ее основная миссия (создание среды для формирования цепочек приращения ценности научной информации). **Заключение.** Результатом реализации предложенной Стратегии должны стать новые высокоинтеллектуальные сервисы для информационной поддержки различных видов научной деятельности; развитие цифровых компетенций исследователей; значительное увеличение интенсивности использования научного контента; развитие рынка научно-информационных услуг; улучшения в системе разделения научного труда.

Ключевые слова: наука о данных, научная информация, научно-техническая информация, онтология, метаданные, научная коммуникация

Для цитирования: Национальная научно-информационная инфраструктура: проблемы, задачи и перспективы / А. Е. Гуськов [и др.] // Управление наукой и наукометрия. 2022. Т. 17, № 3. С. 380—407. DOI: https://doi.org/10.33873/2686-6706.2022.17-3.380-407

Благодарности: статья подготовлена на основе материалов проекта, выполненного в рамках Школы управления исследовательскими программами МШУ «Сколково». Авторы выражают признательность организаторам и модераторам Школы за помощь и наставничество при подготовке проекта, а также профессору РАН Оганову Артему Ромаевичу и д. э. н. Ильиной Ирине Евгеньевне за поддержку при обсуждении текста статьи.

National Scientific Information Infrastructure: Problems, Tasks and Prospects

Andrey E. Guskov[1]✉, Aleksandr S. Karaush[2], Ilya E. Menshchikov[3], Andrei V. Shkolin[3], Vitaly O. Nedelskiy[4], Denis S. Sabirov[5], Timur N. Shchukin[6]

[1] Russian Research Institute of Economics, Politics and Law in Science and Technology (RIEPL), Moscow, Russia
[2] Russian National Public Library for Science and Technology, Moscow, Russia
[3] A. N. Frumkin Institute of Physical Chemistry and Electrochemistry of the Russian Academy of Sciences (IPCE RAS), Moscow, Russia
[4] Gorno-Altaysk State University, Gorno-Altaisk, Russia
[5] Institute of Petrochemistry and Catalysis, Ufa Federal Research Centre of RAS, Ufa, Russia
[6] Moscow School of Management SKOLKOVO, Moscow Region, Russia
✉ *guskov.andrey@gmail.com*

Abstract

Introduction. Large volumes of scientific data and the development of the Data-Driven Science paradigm create a need for fundamentally new tools with which to work with information and conduct scientific communication. The emergence of such tools is only possible given the prerequisite technological foundation – a well-developed digital infrastructure for the storage and exchange of a large variety of scientific data, intended for use not only by researchers but also by software agents. The article analyzes the problems in the established practices of scientific

information management in Russia and the existing barriers that prevent the creation of a new generation of tools. Overcoming them is a very complex and large-scale task, due to the diversity of actors involved in scientific information, the types of scientific information, the specificity of various subject areas, and other factors. To do so, we need a Strategy for the Development of a National Scientific Information Infrastructure, a set of systemic changes in the standards and practices for working with scientific information which will affect a significant part of the scientific community. **Results and Discussion.** The main provisions of the Strategy are formulated, including the goal, objectives and principles of development, its main mission (creating an environment for the formation of value chains of scientific information) and key results – new highly intelligent services for informational support of various types of scientific activity. **Conclusion.** The implementation of the proposed Strategy should result in new highly intelligent services offering informational support for various types of scientific activity; development of digital competencies among researchers; a significant increase in the use of scientific content; development of the market for scientific information services; improvements in the system of division of scientific labor.

Keywords: data science, scientific information, scientific and technical information, ontology, metadata, scientific communication

For citation: Guskov AE, Karaush AS, Menshchikov IE, Shkolin AV, Nedelskiy VO, Sabirov DS, Shchukin TN. National Scientific Information Infrastructure: Problems, Tasks and Prospects. *Science Governance and Scientometrics*. 2022;17(3):380-407. DOI: https://doi.org/10.33873/2686-6706.2022.17-3.380-407

Acknowledgements: this article was prepared on the basis of materials from a project carried out as part of the School of Management Research Program at the Skolkovo Moscow School of Management. The authors are grateful to the organizers and moderators of the School for their assistance and mentoring in the preparation of the project, as well as to Professor Artem R. Oganov and Doctor of Economics Irina E. Ilina for their support in discussing the text of the article.

Введение / Introduction

Уровень научно-технического развития страны определяет ее способность отвечать на большие вызовы, потенциал экономического развития и конкурентность на мировых рынках. Он зависит от целого ряда факторов, в т. ч. от системы организации исследовательской деятельности, развитости научной инфраструктуры, средств производства знания.

Научно-информационная инфраструктура[1] является одним из критически важных элементов, существенно влияющим на уровень

[1] Здесь и далее под научно-информационной инфраструктурой мы будем понимать совокупность объектов, которые предоставляют исследовательскому сообществу ресурсы и услуги, связанные с поиском, сбором, изучением, обработкой, анализом, синтезом и распространением научной информации. К таким объектам прежде всего относятся организации различных форм собственности, а также каталоги, базы данных, информационные системы и сервисы различной принадлежности.

научно-технологического развития страны, поскольку она выполняет функции сохранения и распространения знаний в научном сообществе, пронизывая все дисциплины и этапы жизненного цикла исследований. Так, этапы поиска, сбора и обработки научно-технической информации во многом определяют постановку задачи, планирование исследования и, следовательно , ценность и значимость его результатов. Пренебрежение этим этапом приводит к использованию устаревших методов, работе за границами актуальной научной повестки, дублированию одних и тех же исследований и другим последствиям низкопродуктивной научной деятельности.

По данным разных исследований, работа с научной информацией (ее поиск, сбор, изучение, обработка, анализ, синтез, подготовка к публикации и распространение) была и остается одним из наиболее трудоемких элементов в системе разделения труда исследователей. По некоторым оценкам только чтение научных статей занимает не менее 10 % времени исследователей, и этот показатель продолжает расти [1]. Многие ученые признают, что вынуждены тратить все больше времени на поиск и изучение научной литературы[2]. При этом некоторые информационные специалисты считают, что многие исследователи уже достигли того максимального лимита времени, которое они могли бы выделять на чтение статей [2]. Это говорит о том, что сложившиеся методы работы с научной информацией становятся «бутылочным горлышком» — принципиальным ограничителем эффективности исследовательской деятельности. На фоне быстрого роста объемов знаний, которые уже распространяются не только в формате традиционных публикаций, но и в виде разнообразных массивов данных, это ограничение уже стало критическим.

Кроме того, следует отметить, что современная наука уже не может базироваться на повестке, сформированной исключительно на основе данных из научных статей. Современные реалии требуют концентрации внимания ученых на технологических, предпринимательских, социальных и других аспектах проводимых исследований, т. е. в общий пул научно-технической информации активно вовлекаются данные из патентов, аналитических отчетов, экономических прогнозов и т. д.; становится все более весомой и ценной междисциплинарная составляющая.

Таким образом, современная культура исследовательской деятельности, основанная на необходимости обработки больших объемов данных, требует новых инструментов работы с информацией, а также специально организованной научно-информационной инфраструктуры, где эти инструменты могут функционировать.

Цель настоящей работы заключается в анализе современных тенденций развития научно-информационной инфраструктуры, а также в поиске возможных векторов ее развития с учетом появления все больших объемов информации и научных данных, несоразмерных возможностям ее обработки стандартными методами и усложняющих постановку актуальной научно-технической повестки исследований.

[2] Fleming N. Shut-in Scientists are Spending More Time on Research Papers // Nature Index. 2020. Available at: https://www.natureindex.com/news-blog/shut-in-scientists-are-spending-more-time-on-research-papers (accessed: 07.07.2022).

Результаты и дискуссия / Results and Discussion

Парадигма Data-Driven Science

В последние 20 лет развитие научно-информационной инфраструктуры в мире шло очень интенсивно, в результате чего сформировался новый рынок ресурсов и сервисов для научных коммуникаций и управления научной информацией на разных этапах жизненного цикла исследований. Исследовательские группы выбирают подходящую комбинацию инструментов, на основании которых организуют свою научную деятельность, а крупные компании, такие как Elsevier и Springer, стремясь увеличить число своих клиентов, разрабатывают целые экосистемы связанных между собой сервисов, внедряя их в мировую научно-информационную инфраструктуру. Количество таких инструментов уже исчисляется сотнями, а для их систематизации создаются специальные навигаторы (рис. 1).

Причиной резкого роста количества таких инструментов является повсеместное *развитие новой парадигмы исследовательской деятельности Data-Driven Science*, основанной на сборе, обработке и использовании данных из различных источников[3]. Это не новая научная дисциплина, а особый междисциплинарный подход, который существенно меняет содержание исследовательской деятельности во всех без исключения научных направлениях.

Текущая четвертая парадигма в науке отличается от трех предыдущих тем, что в ней объединяются эксперимент, теория и моделирование[4]. До нового времени существовала экспериментальная наука, в основе которой лежали наблюдения и эксперименты с систематизацией и обобщением их результатов. С XVII в. начинается 2-й этап — развитие теоретической науки, предполагающей использование теоретических моделей и строгое доказательство научных законов. Появление компьютеров в середине XX в. дало старт третьей парадигме науки — науке вычислительной, основанной на методах численного моделирования. «Четвертая парадигма обеспечивает интегрирующую структуру, которая позволяет первым трем взаимодействовать и усиливать друг друга»[5]. Она не просто расширяет объем данных для вычислительных моделей третьей парадигмы, но и возвращает в фокус внимания главный модус работы второй парадигмы — теоретическую работу. *Data-Driven Science* — это не просто работы с большим, чем раньше, количеством данных. Это еще и работа на стыке и поверх множества дисциплин, относящихся к разным типам науки, — например, в междисциплинарном пространстве гуманитарных, естественных и социальных наук. Это означает, что теоретическая, или, шире, онтологическая работа вновь становится ключевой деятельностью ученого. Сложность этой работы в четвертой парадигме значительно превосходит сложность теоретической

[3] The Fourth Paradigm: Data-Intensive Science Discovery / ed. by T. Hey, S. Tansley, K. Tolle. Redmond, Washington: Microsoft Research, 2009. 254 p.

[4] Talk by Jim Gray to the NRC-CSTB in Mountain View, CA, on January 11, 2007. Available at: http://jimgray.azurewebsites.net/jimgraytalks.htm (accessed: 07.07.2022).

[5] Clifford A. Lynch. Jim Gray's Fourth Paradigm and the Construction of the Scientific Record. In: The Fourth Paradigm: Data-Intensive Scientific Discovery / ed. by T. Hey, S. Tansley, K. Tolle. Redmond, WA: Microsoft Research, 2009. P. 177—183.

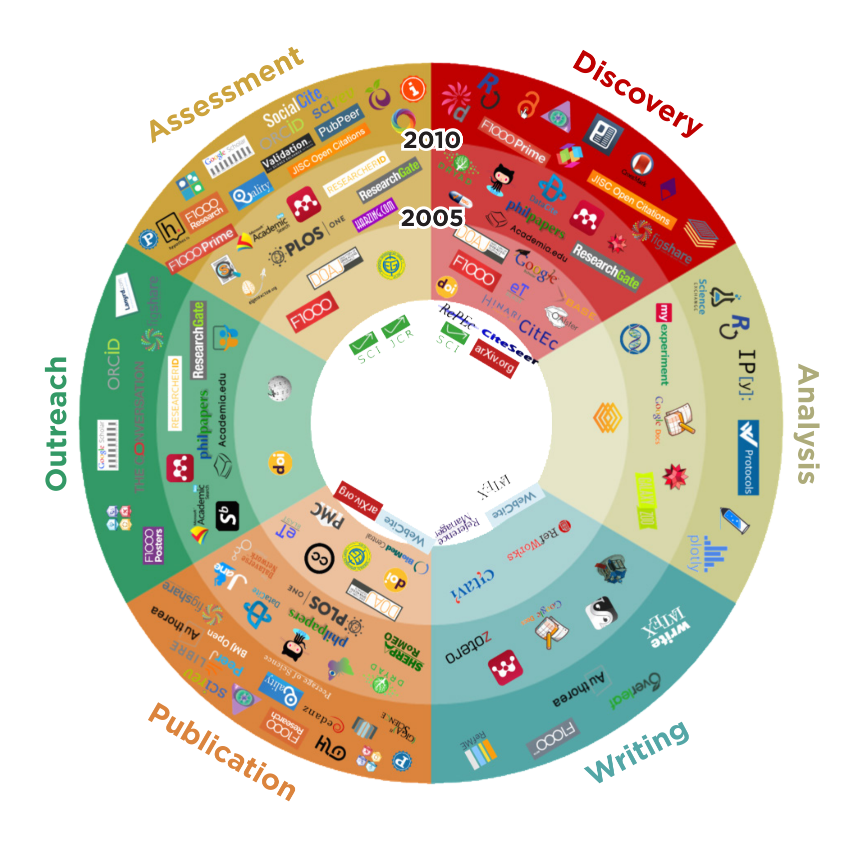


Рис. 1. Наиболее известные инновационные инструменты для научных коммуникаций, распределенные по времени и этапам жизненного цикла исследований

Fig. 1. The most well-known tools for scientific communication, distributed by time and stages in the life cycle of research

Источник / Source: Innovations in Scholarly Communication. Available at: https://101innovations.wordpress.com/press/ (accessed: 07.07.2022).

работы второй парадигмы, а объемы данных и сложность вычислительных моделей могут уже на порядок превышать привычные для третьей, вычислительной парадигмы науки.

Важно отметить, что в исследовательской деятельности происходит еще один фундаментальный сдвиг — трансформация системы разделения научного труда. Так, в третьей парадигме сформировались новые позиции, т. н. «синие воротнички в науке» — специалисты, чья деятельность предполагает хранение, использование и обработку информации. В четвертой парадигме их частично будут замещать новые позиции — «воротнички цифровые» или «цифровые помощники», функции которых лежат в области функций поиска, сбора и анализа научной информации. Это расширит возможности исследователя не только в сфере работы с вычислительными моделями в ситуации порядкового роста количества данных, но и оснастит

его инструментами синтеза данных разного качества и построенных в разных дисциплинарных логиках и различными разметками данных, т. е. в разных дисциплинарных онтологиях[6].

Можно даже сказать, что этот тип инструментов неминуемо обретает функции агентности — находясь в коммуникации и взаимодействии с исследователем, программный агент парадигмы Data-Driven Science будет вынужден демонстрировать элементы самостоятельности в собственной активности, осуществлять проактивный выбор синтетических меж- и трансдисциплинарных моделей, а его взаимодействие с исследователем повысит качество коммуникации и диалога, а не будет просто выполнением его команд. Фактически цифровой помощник все больше становится тем, кто вместо исследователя обрабатывает массивы информации, «читает» сотни и тысячи публикаций, фокусируя внимание на выделенных им наиболее значимых моделях-гипотезах подобно тому, как раньше это делали эксперты институтов научной информации.

Принципиально важно отметить, что агент такого рода не просто осуществляет поиск «подходящих» данных и экспериментальных дизайнов в рамках одной дисциплинарной модели. Работая с представленными в метаданных теоретическими конструктами, он осуществляет трансдисциплинарный поиск и синтез новых гипотез, указывая на скрытое подобие в онтологических конструкциях разных дисциплин. Цифровой агент может предложить перенос и достройку элементов этих конструкций из одной дисциплины в другую, а также онтологические основания, способные объединять результаты работы исследователей из разных дисциплин. Примером такой работы, осуществляемой сегодня вручную, являются онтологии, воплощенные в платформенные и сетевые научно-исследовательские сервисы: геномный браузер, лингвистическая платформа Lingvodoc и др.

Программные агенты выполняют рутинные действия с данными быстрее людей, однако гораздо хуже «настраиваются» на различные формы их представления. Поэтому их появление формирует новые *требования* к научно-информационной инфраструктуре:

- доступность качественной метаинформации о научных результатах;
- унификация форматов и протоколов обмена данными;
- формализация предметных областей.

Очевидно, что полная унификация форматов и протоколов данных является недостижимой целью, а требуемая для этого унификация онтологических оснований различных дисциплин принципиально невозможна (согласно второй теореме Геделя). Однако обеспечение автоматизированного доступа программным агентам к метаданным результатов научной деятельности потребует унификации и перевода в машиночитаемый вид методологических элементов этих результатов — экспериментальных планов, теоретических оснований, используемой рабочей онтологии и т. д.

Поддержка этих требований значительно повышает видимость результатов научных исследований и обеспечивает возможность

[6] Под онтологией в этой статье мы будем понимать формальное описание некоторой области знаний, которое, как правило, содержит все релевантные классы объектов, связи между ними и правила, принятые в этой области.

создания нового поколения сервисов обработки научной информации, основанных на анализе больших данных методами искусственного интеллекта. Их невыполнение сохраняет принципиальные барьеры развитию Data-Driven Science, ограничивает производительность труда исследователей и снижает конкурентоспособность отечественной науки в мире, где уже совершается переход к четвертой парадигме.

Проблемы российской научно-информационной инфраструктуры

Проблема несоразмерности объема научной информации возможностям человека появилась еще в прошлом веке, когда счет издаваемых журналов пошел на сотни и тысячи. Ответом на этот вызов стало создание специализированных организаций — институтов научной информации (ВИНИТИ и ИНИОН в СССР, ISI в США), основными задачами которых были сбор и хранение научных материалов, полученных из разных стран мира, а также создание реферативных журналов, посредством которых можно было бегло ознакомиться с результатами последних научных исследований. Следует отметить различную судьбу этих организаций в 1990-х гг.: пока российские институты вместе со страной переживали глубокий кризис, американский ISI со своим продуктом Web of Science стал мировым лидером в области библиографических баз данных, перейдя на предоставление услуг в формате онлайн. С тех пор и до настоящего времени его основной продукт является одним из ключевых элементов научно-информационной инфраструктуры многих государств и мировой науки в целом.

Отметим, что вопросы развития научно-информационной инфраструктуры обычно являются одним из приоритетов государственной научной политики. Например, в Европейском союзе создали специализированную правовую форму, которая называется Европейский консорциум научной инфраструктуры (European Research Infrastructure Consortium). В рамках этих консорциумов реализуются различные проекты [3], за счет которых развитие европейской научно-информационной инфраструктуры идет очень интенсивно.

По ряду причин научно-информационная инфраструктура в России на настоящий момент слабо ориентирована на поддержку исследований в парадигме Data-Driven Science. Созданный в 1970-х гг. мощный задел в виде Государственной системы научно-технической информации (ГСНТИ) давно утрачен, а существующая инфраструктура с 1990-х гг. формировалась стихийно, практически без централизованных решений. Попытки их создания предпринимались в институтах РАН[7], однако они не получили широкого распространения (исключениями являются Российский индекс научного цитирования (далее — РИНЦ) и несколько нишевых проектов, таких как www.mathnet.ru). Значительную роль в этом сыграли недостаточное финансирование (до 2012 г. научной сфере в целом не уделялось должного внимания, и в частности — вопросам управления научной информацией) и отсутствие систематической реализации программного подхода к проблеме (стратегические программы не разрабатывались и не реализовывались). Это привело к т. н. «лоскутной

[7] Кулагин М. В., В. А. Серебряков. Информационное пространство РАН (Проекты и реализация, 1998—2013) // Тр. XVIII Всерос. науч. конф. Новороссийск, 19—24 сентября 2016 г. Новороссийск: Ин-т приклад. математики им. М. В. Келдыша РАН, 2016. С. 194—221.

автоматизации» в масштабах страны — разработке отдельных решений (чаще государственных и реже коммерческих), слабо пригодных для взаимной интеграции и для формирования эффективной системы распространения знаний, как среди российских ученых, так и их интеграции в мировую науку. Не были сформулированы единые принципы и не была разработана (или хотя бы заимствована) единая модель целевого состояния системы научно-технической информации. Будь это сделано, принцип обеспечения машиночитаемости научно-технических данных программными агентами и соответствующая этой практике нормативная база были бы сформулированы в качестве первых требований.

Деятельность институтов научной информации и научных библиотек атомизирована, их программы развития никак не связаны между собой, а нередко просто отсутствуют. Ситуация усугубляется из-за неразвитого национального рынка научно-информационных продуктов и, как следствие, высокой зависимости от информационных ресурсов иностранного производства.

На государственном уровне обсуждалось создание централизованного решения — Единого российского электронного пространства знаний, упомянутого в Федеральном законе «О библиотечном деле» (ред. от 03.07.2016). Однако этот проект пока не реализован, и более того, его концепция также не утверждена. По нашему мнению, основным препятствием для его использования в научной сфере является ведомственная принадлежность. В подзаконных актах сказано, что «Пространство знаний формируется Министерством культуры Российской Федерации на технологической основе Национальной электронной библиотеки»[8], тогда как большинство поставщиков и потребителей научной информации относятся к Минобрнауки России, а также к другим ведомствам и госкорпорациям, занимающимся исследованиями и разработками. Это создает высокие риски, что такая система будет ориентирована, в первую очередь, на сохранение исторического и культурного достояния народов России, а обсуждаемые здесь проблемы научно-информационной инфраструктуры так и не будут решены.

Вопросам развития Национальной научно-информационной инфраструктуры не уделялось должного внимания еще и из-за того, что существует *иллюзия доступности научной информации*. Она заключается в том, что для опытного исследователя не составит труда найти нужную публикацию в одном из многих возможных каналов: РИНЦ, лицензионная подписка организации, университетские архивы, репозитории препринтов, научные социальные сети, обращения к коллегам из других организаций, разные виды открытого доступа. Иллюзорность такой доступности состоит в том, что этот подход нельзя «поставить на поток», алгоритмизировать и запрограммировать. Это создает непреодолимый барьер для активного включения в научную коммуникацию программных агентов и принципиально ограничивает переход к следующей парадигме науки в стране,

[8] Постановление Правительства РФ от 20.02.2019 № 169 «Об утверждении Положения о федеральной государственной информационной системе “Национальная электронная библиотека” и методики отбора объектов Национальной электронной библиотеки» // СПС «КонсультантПлюс».

создавая потолок продуктивности научно-исследовательского и технолого-производственного комплексов страны.

Поэтому перечисленные ниже проблемы мы рассматриваем с позиции архитектора информационных программных систем, задачей которого является выстраивание архитектуры интенсивного информационного обмена, цепочек обработки, сопоставления, систематизации и анализа данных с целью увеличения их информационной ценности в рамках парадигмы Data-Driven Science.

1. Недоступность или низкое качество метаданных[9]. Метаданные публикаций являются базовым информационным элементом любой информационной системы. Их плохое качество снижает функциональность таких систем, а отсутствие делает невозможным их создание и развитие[10].

Основные массивы метаданных российских научных публикаций агрегируются в системе РИНЦ, доступ к который открыт программным агентам только на коммерческой основе. Кроме РИНЦ, эти данные могут поступать в один из центров регистрации идентификаторов DOI (как правило, это система Crossref). Однако зачастую редакции журналов передают информацию несвоевременно, с несоблюдением части требований к размещаемой информации.

2. Доступ к полным текстам публикаций затруднен. Развитие парадигмы Data-Driven Science основано и реализуемо только при условии открытости научной информации и унификации протоколов доступа к ней для потоковой машинной обработки. Однако сегодня открытые централизованные полнотекстовые архивы в России отсутствуют за очень редкими исключениями (Cyberleninka, Mathnet), более того, многие оттиски публикаций находятся в зашифрованном виде с запретом чтения и копирования. Электронно-библиотечные системы предоставляют лицензированный доступ к своим ресурсам и при этом сфокусированы на задаче книгообеспеченности образовательного процесса в вузах.

В этом смысле агрегаторы научной информации (например, РИНЦ) могут обеспечивать не только пользовательский доступ к публикациям в человеко-читаемом виде и стандартным метаданным (названия, авторы, даты), но и обеспечивать возможность их углубленной обработки через программные интерфейсы.

3. Результаты научных экспериментов (в т. ч. дорогостоящих), на основании которых публикуются научные работы, остаются недоступными. Во многих случаях они некоторое время хранятся на жестких дисках лабораторных компьютеров и затем часто пропадают. Из-за этого полученные результаты анализа невозможно проверить, а первичные данные — повторно использовать в других

[9] Метаданные — это данные о данных; информация о другой информации или данные, относящиеся к дополнительной информации о содержимом или объекте. Например, метаданными статьи является ее название, наименование журнала, коды тематических классификаторов, ФИО авторов, места их работы и др. Метаданные имеют строгую структуру и позволяет обрабатывать объект, не извлекая его содержимое (например, не анализируя полный текст публикации).

[10] Гуськов А. Е., Косяков Д. В. С чего должна начинаться цифровизация российской науки? URL: https://indicator.ru/humanitarian-science/s-chego-dolzhna-nachinatsya-cifrovizaciya-rossiiskoi-nauki.htm (дата обращения: 07.07.2022).

исследованиях. Отметим, что в современной исследовательской культуре массивы данных уже стали самостоятельным видом публикаций, издаваемых в специальных высокорейтинговых журналах, например, Scientific Data (Nature Publishing Group), Data in Brief (Elsevier). В России подобная практика отсутствует.

4. Ряд профинансированных проектов в области развития научно-информационной инфраструктуры не дали ожидаемого эффекта. В качестве некоторых примеров можно привести проект ScienceID, который разрабатывается с 2017 г. и до сих пор не стал частью информационной инфраструктуры, а также государственные конкурсы на создание тематических архивов в 2020 г., большинство из которых либо не работают, либо не используются. Причиной является недостаток механизмов поддержки уже начатых проектов, отсутствие «зонтичной» программы, которая бы «втягивала» в себя новые перспективные решения, интегрировала бы их в существующую инфраструктуру и оказывала содействие в их развитии.

Эти проблемы ясно говорят о серьезных трудностях, возникающих при создании высокотехнологичных информационных продуктов на базе существующей в нынешнем виде отечественной инфраструктуры. В данном случае речь идет не только о коммерческих сервисах, но и об исследовательских платформенных решениях, вокруг которых формируются профессиональные сообщества предметных специалистов.

Принципы разработки Стратегии развития Национальной научно-информационной инфраструктуры

Необходимо отметить, что Национальная научно-информационная инфраструктура является очень сложным объектом управления. Эта сложность обусловлена разнообразием акторов научно-информационной деятельности (поставщики и потребители научной информации, ее систематизаторы, владельцы государственных и коммерческих информационных систем), видов научной информации (монографии, статьи, диссертации, патенты, отчеты, датасеты, модели и др.), спецификой различных предметных областей и другими факторами. Перечисленные выше проблемы имеют системный характер, а значит, и их решение должно представлять собой комплекс системных изменений, затрагивающих значительную часть научного сообщества. Поэтому их последствия должны быть детально проанализированы, а сами изменения должны сопровождаться разъяснительными мероприятиями и мерами, стимулирующими распространение новых практик.

Другими словами, существуют принципиальные ограничения на возможность постепенной перестройки системы, особенно в ситуации сильно ограниченного *рынка* научно-технической информации. Мы считаем, что в текущей ситуации введение новых принципов организации комплекса научно-исследовательской информации не может быть отдано на откуп процессам самоорганизации и должно быть инициировано сверху. Причиной невозможности реализации решения, основанного на самоорганизации, которое в любых других условиях было бы наиболее логичным и продуктивным, повторим, является практически полное отсутствие предпринимательской

инициативы в этой зоне деятельности и, соответственно, отсутствие рыночных агентов, способных на самоорганизацию.

Решение обозначенных проблем должно начинаться с разработки Стратегии развития Национальной научно-информационной инфраструктуры, в которой научным сообществом формулируется консенсус целей, задач и направлений ее развития. Создание Стратегии, отраженной в форме комплекса документов и разделяемой ключевыми акторами этой сферы, само по себе является сложной организационной задачей, успешное решение которой возможно лишь при соблюдении принципов коллегиальности, структурности и иерархии целеполагания, а также программного подхода.

Коллегиальность. Основная сложность разработки Стратегии состоит в необходимости учитывать интересы большого числа разнородных участников информационного обмена: поставщиков, потребителей, агрегаторов, систематизаторов, операторов государственных информационных систем и регуляторов (подробнее о них будет сказано ниже), поэтому разработку Стратегии необходимо осуществлять на коллегиальной основе, привлекая представителей этих участников. Этот же механизм обеспечивает распространение положений Стратегии, ее понимание и принятие в научном сообществе. Несоблюдение принципа коллегиальности приведет к плохо продуманным или дискриминирующим решениям, делающим невозможным достижение стратегических целей.

Структурность. Стратегия должна охватывать разные аспекты управления научной информацией, поэтому ее целесообразно разбить на два уровня. На 1-м уровне создается рамочный документ «Концепция Национальной научно-информационной инфраструктуры», определяющий основные цели и задачи, принципы функционирования и развития, ключевых участников и другие общие положения. На 2-м уровне, в соответствии с общей концепцией, разрабатываются стратегии по каждому из ключевых функциональных аспектов, таких как:

1) управление российской научной информацией, включая системы учета, хранения и распространения публикаций, форматов и протоколов обмена метаданными, политики развития открытых репозиториев и управления исследовательскими данными, вопросы развития основных видов научных публикаций, таких как монографии и научная периодика;

2) доступ к зарубежной научной информации, включая декларацию принципов и механизмов реализации национальной и централизованной подписки на полнотекстовые ресурсы, научно-информационные предметные и аналитические базы данных, а также принципы участия в движении Open Access и сопутствующие меры поддержки;

3) развитие предметных онтологий, включая разработку словарей, тезаурусов, таксономий в различных предметных областях, формирование связей между ними, а также разработка механизмов их распространения и массового использования в процессах обмена научной информацией. Разработка методов извлечения научных знаний, их сопоставления, обобщения и анализа;

4) развитие рынка научно-информационных услуг, с учетом разных функциональных сегментов: поиск научной информации,

аналитика, создание предметных баз знаний, высокоинтеллектуальные сервисы обработки научной информации и др.

Этот список аспектов не является исчерпывающим. Более того, может оказаться целесообразным выделение 3-го уровня, где в рамках широких аспектов 2-го уровня будут сформулированы узкие стратегии для более специфичных функциональных аспектов. К ним могут быть отнесены наиболее критичные или сложные вопросы научно-информационной инфраструктуры, например, развитие научной периодики, управление исследовательскими данными, создание научного задела для высокоинтеллектуальной обработки текстов на естественном языке.

Иерархия целеполагания. Перечисленные выше стратегические документы должны определять цели, риски, возможности и ограничения для остальных проектов и программ, направленных на поддержку функционирования и развития Национальной научно-информационной инфраструктуры. Это означает, что на их основе должны разрабатываться документы управления элементами научно-информационной инфраструктуры, в которых определяются конкретные механизмы реализации положений Стратегии (рис. 2). Например, к ним относятся:

— программы развития институтов научной информации, научных библиотек и других организаций, являющихся ключевыми участниками Национальной научно-информационной инфраструктуры;

— технические задания на развитие государственных систем в области управления научной информацией;

— конкурсная документация на выделение грантов для исследований в области развития методов искусственного интеллекта;

— рекомендации по развитию научных журналов, научных издательств и электронно-библиотечных систем;

— рекомендации по управлению научной информацией для университетов и научных организаций;

— профессиональный стандарт для обучения «инфодайверов» — современных специалистов в области поиска, извлечения и обработки научной информации.

Наконец, **программный подход** должен преобладать над проектным, поскольку целью реализации Стратегии является перестройка научно-информационной инфраструктуры с длительным горизонтом планирования. Практика показывает, что проектный подход несет в себе риски краткосрочности эффектов и исключение из фокуса внимания долгосрочных целей. Программный же подход обеспечивает лучшие условия для непрерывного и последовательного развития инфраструктуры, увеличения и улучшения качества связей между ее элементами, а также создает среду и стимулы для появления новых элементов и совершенствования уже существующих. Необходимо отметить, что он является приоритетным и не отменяет проектную деятельность, а определяет для нее рамки ограничений и возможностей. В дальнейшем мы будем называть комплекс мероприятий по развитию Национальной научно-информационной инфраструктуры Программой.

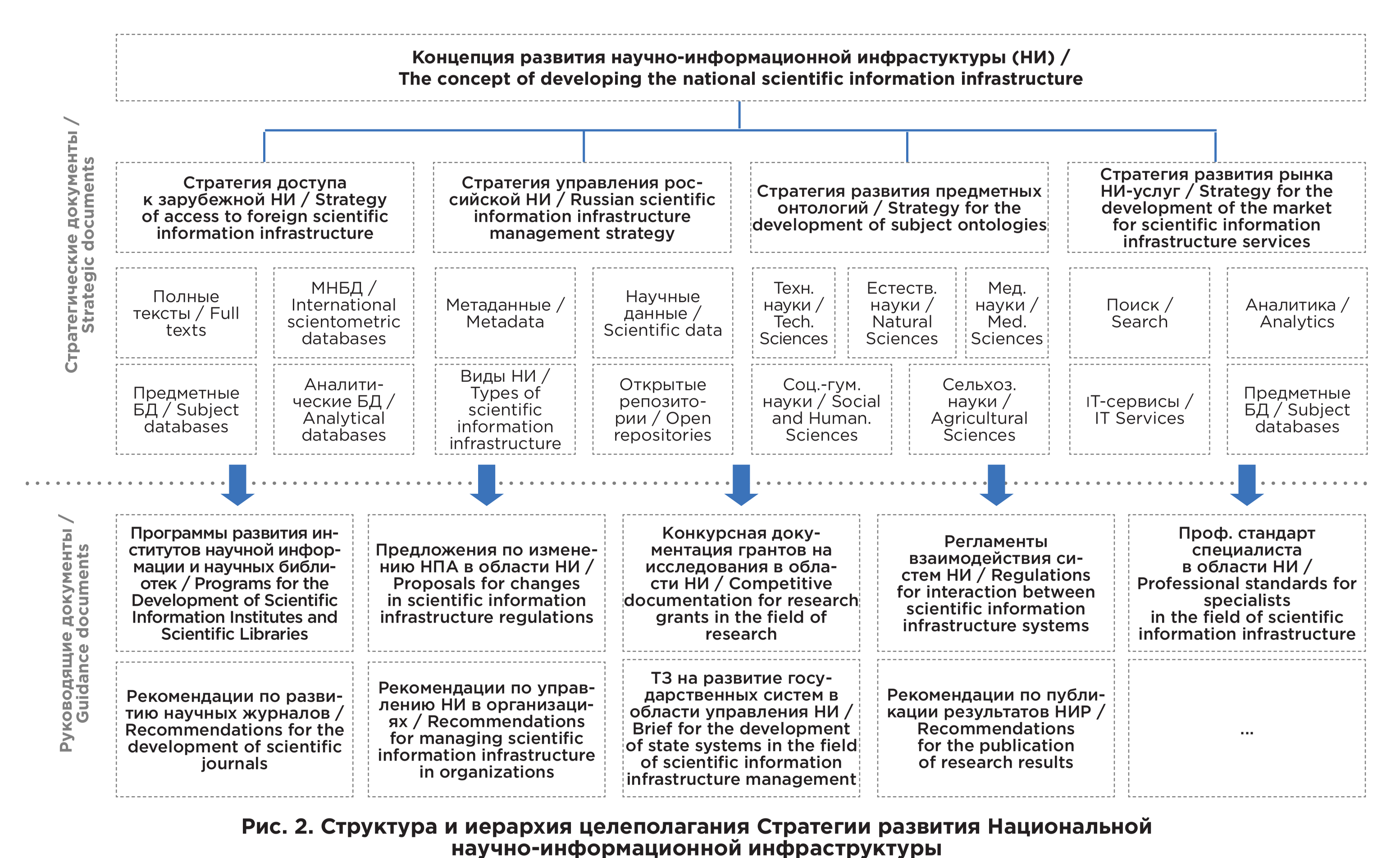


Рис. 2. Структура и иерархия целеполагания Стратегии развития Национальной научно-информационной инфраструктуры
Fig. 2. Structure and hierarchy of goal setting of the National Scientific Information Infrastructure Development Strategy

Цели, задачи и барьеры развития Национальной научно-информационной инфраструктуры

Авторы этой работы считают, что основной целью должно быть развитие инфраструктуры рынка научной информации в соответствии с парадигмой Data-Driven Science. Это означает, что Стратегия должна быть направлена на запуск рыночных механизмов в этой сфере и самоорганизацию агентов разного типа — управленцев, предпринимателей, разработчиков, исследователей и т. д. На наш взгляд, реализация программы в рамках преимущественно бюджетного финансирования не приведет к появлению живой, гибкой и продуктивной системы поддержки Data-Driven Science в стране, в то время как инициация трансформации сверху вкупе с формированием принципов и условий, обеспечивающих возникновение рынка и формирование в нем устойчиво растущей наукоемкой экономики, способны решить поставленную нами задачу.

Фокусировка на новой парадигме создает стимул для развития цифровых инструментов и соответствующих компетенций исследователей, что в конечном итоге повышает конкурентоспособность российской науки. Акцент на коммерческом сегменте будет способствовать притоку инвестиций в эту сферу, стимулирующих появление новых продуктов и повышающих устойчивость всей системы. При этом сама программа должна обеспечивать государственную поддержку развития нижнего уровня инфраструктуры, на котором обеспечивается создание и распространение научной информации. Это необходимое условие, поскольку, во-первых, основная часть научной информации в России создается в государственных организациях преимущественно с бюджетным финансированием, а во-вторых, без необходимых для трансформации государственных инвестиций и прозрачной научно-информационной политики коммерческие компании едва ли будут активно вкладываться в это направление.

При этом инфраструктура должна быть принципиально открытой для других, наиболее продуктивных и активных акторов научно-исследовательского профиля — RnD центров частных компаний, свободных ученых, а также тех позиций, которые традиционно не рассматривались в качестве создателей научного знания — предпринимателей, производственников, разработчиков, представителей свободных творческих профессий, педагогов и других позиций, чье функционирование связано с созданием или потреблением научных знаний. Критически важно как можно раньше сделать инфраструктуру доступной для ученых и предпринимателей из других стран, так ее мощность и продуктивность будет значительно повышена, и у страны появится принципиальная возможность создания нового сектора глобально-конкурентной экономики.

Для достижения этой цели мы выделили 3 ключевые задачи и соответствующие им барьеры, которые придется преодолевать.

Задача 1. Создание распределенной инфраструктуры для сбора, хранения, обработки и распространения обмена и доступа к массивам научной информации в машиночитаемой форме. Эта задача включает создание и развитие научно-информационных систем, обеспечивающих сбор, хранение, обработку и распространение различных видов и форм научных данных, разработку и внедрение

необходимых для этого стандартов и протоколов, создание программных интерфейсов (API) для «цифровых агентов».

Для решения этой задачи необходимо преодоление следующих организационных и технических барьеров.

1. Большой объем существующих и новых научных данных создает сильный инерционный фактор и не позволяет оперативно вносить изменения во все элементы инфраструктуры.

2. Организационная разобщенность владельцев научных данных делает невозможным формирование прямых контуров управления инфраструктурой.

3. «Монополия» на данные, когда владельцы научной информации не заинтересованы в ее распространении, требует введения сложной системы мотивации для внедрения изменений.

4. Российские издатели преимущественно ориентированы на потребление научного контента людьми, а не программами.

5. Недостаточно массовое использование стандартных форматов представления данных и схем идентификации информационных сущностей (публикаций, людей, организаций, проектов), в т. ч. из-за их отсутствия или незнания специалистов об их существовании, значительно затрудняют связывание данных из различных источников.

Задача 2. Создание научного задела для разработки высокотехнологичных продуктов. Эта задача подразумевает стимулирование научных проектов в области развития методов анализа текстов на естественных языках (NLP, Natural Language Processing) с целью извлечения знаний, методов сбора, систематизации и управления знаниями, включая методы искусственного интеллекта и обработки больших данных. Для стимулирования таких проектов целесообразно выделить отдельную исследовательскую программу на базе профильных научных организаций и университетов, а также проводить конкурсы по выделению грантов на инициативные исследования в этой области.

Эта задача нацелена на получение результатов в области т. н. фронтиров — тематик исследований, которые находятся на передовом крае мировой научной мысли. Поэтому барьеры для этой задачи имеют научно-технологический характер.

1. Высокая сложность задачи извлечения знаний из данных и текстов не позволяет рассчитывать на получение быстрых и качественных решений. Первые прототипы будут иметь низкую точность и эффективность, которые со временем будут повышаться.

2. Различные школы и подходы к систематизации знаний затрудняют совместное использование этих результатов. Необходимо предпринимать особые усилия для синхронизации этой деятельности, а также для разработки конвертеров знаний между различными системами, форматами и онтологическими моделями.

Задача 3. Создание системы мотивации для всех пользователей инфраструктуры. Эта задача направлена на развитие рынка научно-информационных услуг. Для ее решения необходимы изучение явных и скрытых потребностей потребителей научной информации, оценка потенциала российского сегмента рынка и поиск ниш на мировом рынке научной информации. Можно сказать, что суть

этой задачи состоит в таком изменении системы разделения труда ученых, при котором часть функций, связанных со сбором, анализом и управлением научной информацией, передается от исследователя к цифровым помощникам, которые рано или поздно смогут выполнять эти функции и быстрее, и дешевле, и, возможно, качественнее.

К основным барьерам для этой задачи следует отнести следующие:

1) сложившаяся во многих российских научных организациях система разделения труда ученых, которая сформировалась еще в прошлом веке и с тех пор почти не менялась, слабо ориентирована на адаптацию к парадигме Data-Driven Science и использование информационных продуктов нового поколения;

2) неразвитость российского рынка научно-информационных услуг ограничивает рост исследовательской культуры и притупляет информационные потребности ученых;

3) недостаток в научном секторе квалифицированных специалистов в области работы с научной информацией (да и в области информационных технологий вообще) ограничивает проникновение новых продуктов в исследовательские институты и университеты.

Национальная инфраструктура как среда для приращения ценности научной информации

Поясним суть 1-й и 2-й задач с помощью процесса приращения ценности научной информации. На рис. 3 изображены 4 уровня ценности научной информации с точки зрения использующих ее продуктов. Если на нижнем уровне эта ценность состоит лишь в том, что отдельно взятая публикация может быть кем-то получена и прочитана, то на 2-м уровне подобные публикации с помощью каталогов уже формируют организованное информационное пространство, в котором значительно упрощается их поиск и разные виды потоковой обработки. На 3-м уровне на основе извлеченных из публикаций данных формируются базы знаний — продукты, ключевым свойством которых является возможность машинного «понимания» и «оперирования» знаниями, представленными некоторым формализованным способом. На самом высокотехнологичном, 4-м, уровне создаются интеллектуальные программные сервисы, которые могут не только анализировать и обобщать имеющиеся формализованные знания, но и порождать новые, производные от них.

Необходимо отдельно отметить иерархичность представленных уровней приращения ценности научной информации. Во-первых, уровни 1 и 2 соответствуют описанной выше задаче 1, а уровни 3 и 4 — задаче 2. Во-вторых, нижестоящие уровни не только предоставляют пользователям набор самостоятельных продуктов, но и формируют из них функциональную основу для продуктов вышестоящих уровней. Продукты верхних уровней используют эту основу, к которой применяют новые методы сбора, структуризации, извлечения, обогащения или анализа научного знания, что обеспечивает решение все более трудных задач.

Ценность решения этих задач состоит в том, что трудозатраты исследователей на работу с научной информацией сокращаются за счет частичного или полного делегирования этих задач «цифровым помощникам». Действительно, на 1-м уровне все задачи выполняет

Виды продуктов / Types of products	Уровни обработки научной информации / Levels of scientific information processing	Новые свойства / New properties	Барьеры / Barriers
4. Интеллектуальные сервисы / Intelligent services	Анализ обобщение и генерация знаний / Analysis, synthesis, and generation of knowledge	Машинное создание производных (вторичных) знаний / Machine creation of derived (secondary) knowledge	Научный фронтир / Scientific Frontier
3. Базы знаний / Knowledge bases	Извлечение и сопоставление знаний / Extracting and comparing knowledge	Машинное «понимание» знаний / Machine "understanding" of knowledge	Междисциплинарная рассогласованность, сложность извлечения знаний / Interdisciplinary incoherence, difficulty in extracting knowledge
2. Каталоги научных публикаций / Catalogs of scientific publications	Агрегирование и структурирование массивов научной информации / Aggregating and structuring arrays of scientific information	Потоковая (машинная) обработка, поиск / Streaming (machine) processing, search	Большой объем, рассогласованность форматов / Large volume, format inconsistency
1. Доступ к публикациям / Access to publications	Хранение и предоставление доступа / Storing and providing access	Сохранность, доступность / Preservation, accessibility	Организационная разобщенность, «монополия» на данные / Organizational fragmentation, a data "monopoly

Приращение ценности научной информации / Increasing the value of scientific information

Рис. 3. Уровни приращения ценности научной информации
Fig. 3. Levels of incremental value of scientific information

сам исследователь, а возможности для автоматизации отсутствуют или очень скупы. На 2-м уровне программные агенты получают возможность работать с каталогами научных публикаций, в значительной степени сокращая для ученых время на поиск и выборку нужных им работ. На 3-м уровне «машины» помогают исследователям «читать»: анализируют наборы данных, модели, методы и другие элементы, выявляют ключевые сущности (материалы, реактивы, препараты, оборудование), основные тезисы и извлекают их содержание в виде формализованных моделей. Эти модели используются на 4-м уровне, где на их основе интеллектуальные сервисы помогают ученым создавать производные знания.

Таким образом, основной миссией Программы является создание «питательной среды» для формирования цепочек приращения ценности научной информации, результатом чего будут являться высокоинтеллектуальные сервисы для информационной поддержки различных видов научной деятельности. Как уже говорилось, их массовое появление и успешное развитие как продуктов возможно лишь в условиях рынка с хорошо выраженным спросом. Какого рода продукты это могут быть?

Приведем здесь пример принципиальной архитектуры и ориентировочный сценарий работы научно-информационного сервиса нового поколения «Навигатор для планирования исследований» (рис. 4), появление которого в той или иной реализации представляется

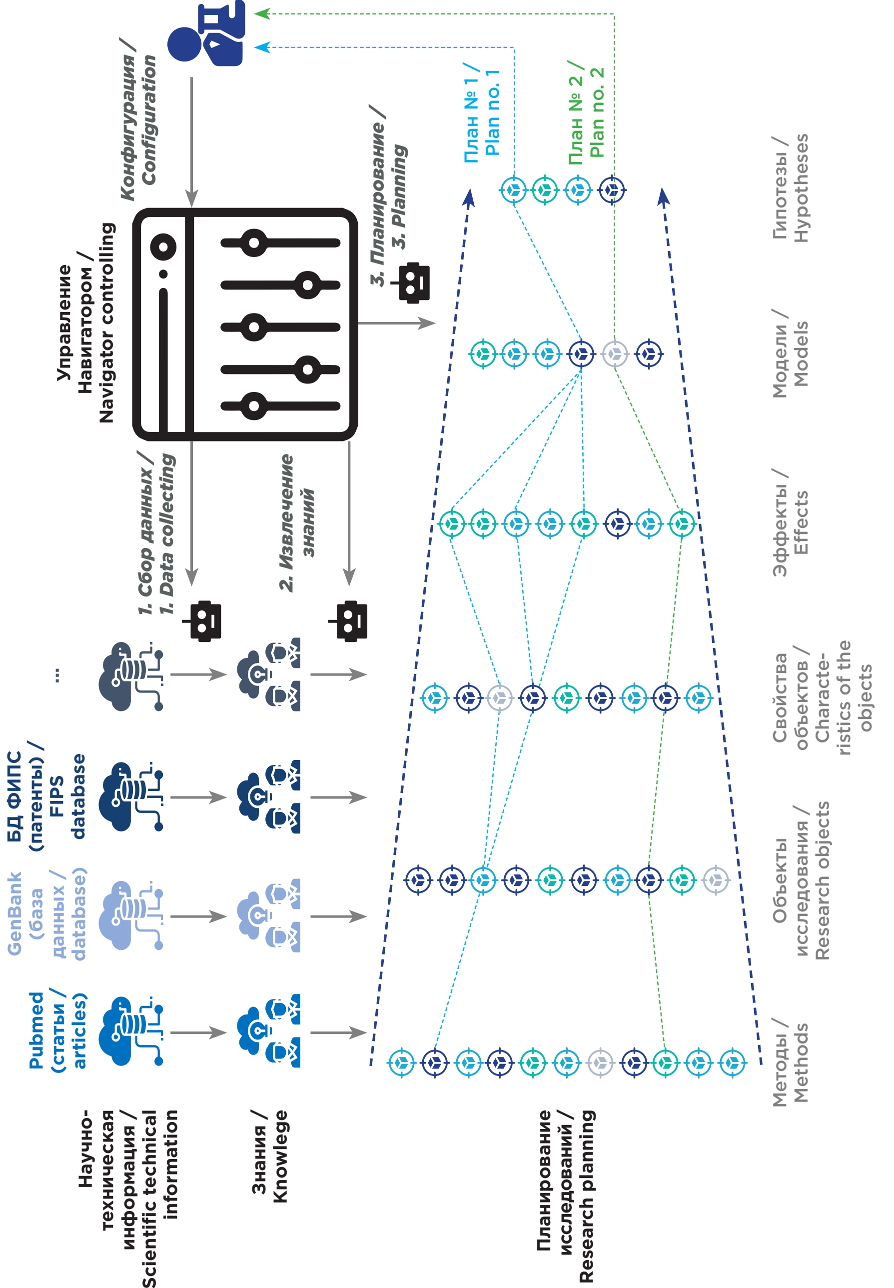


Рис. 4. Схема функционирования сервиса «Навигатор для планирования исследований»
Fig. 4. Functional scheme of the service "Navigator for research planning"

неизбежным на горизонте 5—10 лет как минимум в области естественных и медицинских наук, после чего эти принципы будут реализованы на более широком дисциплинарном спектре.

В начале работы исследователь конфигурирует требования к работе агента в соответствии с масштабом своей научной задачи, описывая спецификацию одного или нескольких исследовательских вопросов, сформулированных в его проекте, или обозначая те характеристики будущего исследования, которые задают рамки работы программного агента — например, методы, применимость, ограничения оборудования и ресурсов, требования к глубине коллаборации, дисциплинарный набор. Чем уже эти рамки, тем меньше потенциально интересных направлений исследований будет найдено, но тем проще и быстрее будет выполнена работа. В зависимости от исходной конфигурации могут выполняться задачи от поиска альтернативных экспериментальных моделей или материалов под уже определенное исследование вплоть до широкого поиска фронтиров с высоким теоретическим или практическим потенциалом в широком междисциплинарном поле.

Навигатор обращается во внешние источники данных и собирает из них научную информацию, которая может быть релевантна для этой задачи: статьи, патенты, датасеты и др. Затем из этой информации Навигатор извлекает элементы знаний и связи между ними: объекты исследований и их свойства, методы, модели и др. При работе в разных дисциплинарных доменах агент строит собственную онтологическую карту проблемной области, указывая на совпадения и различия в теоретических основаниях, стоящих за экспериментальными моделями, и предлагает исследователю выбрать тот или иной набор теоретических и методологических оснований или помогает ему сконфигурировать новую онтологию и, таким образом, новый объект исследований.

С помощью построенной таким образом онтологической модели Навигатор формирует варианты планов исследований с указанием необходимых условий, например наличия определенного оборудования и реактивов. Кроме того, Навигатор указывает на необходимые коррекции в онтологических и методологических основаниях исследования, которые делают возможным привлечение датасетов, экспериментального дизайна или понятийного ряда из смежных или даже далеких дисциплин.

Для каждого элемента плана приводится обоснование его включения в план с указанием источников, на основании которых это решение было предложено. Исследователь, исходя из имеющихся в его распоряжении возможностей, выбирает соответствующий план и запускает исследовательский процесс. Данный пример не описывает полный цикл исследования, и по этой причине мы не рассматриваем то, как организован экспериментальный и вычислительный этапы исследовательской работы. Достаточно сказать, что эти этапы также должны быть оснащены соответствующим инструментарием, который позволит использовать их результаты в последующей работе этой или любой другой исследовательской группы. К таким инструментам относятся, например, системы цифровых лабораторных журналов и другие системы, обеспечивающие интеграцию лаборатории в цифровую инфраструктуру четвертой парадигмы науки.

Другим примером продуктов нового поколения является сервис подготовки научных обзоров, который находит публикации по заданной тематике за определенный период времени, анализирует их содержимое, производит обобщение и формирует текст обзора публикаций. Отметим, что этот текст будет не полноценной статьей, а, скорее, «полуфабрикатом» для локального применения. Чаще всего он будет использоваться для быстрого ознакомления исследователей с последними результатами, полученными другими научными группами. Вместо изучения десятков и сотен статей достаточно будет ознакомиться с одним обзором-«полуфабрикатом», выбрав на его основе несколько работ, которые ученому следует внимательно прочитать.

Также обзор-«полуфабрикат» может выступить прообразом настоящего научного обзора, если группа исследователей использует его в качестве основы и доработает до полноценной публикации. Отметим, что по мере развития методов обработки текстов на естественном языке его качество будет постепенно возрастать и требовать все меньше усилий по его доработке. Рано или поздно цифровой помощник самостоятельно сможет сгенерировать полноценный научный обзор, и это уже не является фантастикой — первая монография о технологиях создания литий-ионных батареек, созданная программными алгоритмами, была издана в 2019 г. [4].

Может показаться, что в условиях развитой научно-информационной инфраструктуры у экспертов в области научной информации работы будет значительно меньше. Полагаем, что все будет ровно наоборот — задач, связанных с обработкой научной информации, станет значительно больше, поскольку исследователям будет выгодно не тратить на это свое время, а воспользоваться высококвалифицированным сервисом. При этом деятельность информационных специалистов уже в меньшей степени будет рутинной (такие задачи будут выполняться программными средствами), а в большей — интеллектуальной, включая совершенствование методов машинного обучения.

В скором времени, вероятно, начнется процесс «уберизации» рынка услуг научной информации. Будут появляться «маркетплейсы» научных инфосервисов, которые будут упрощать связывание поставщиков типовых услуг (отдельных информационных специалистов и компании, которые специализируются в области обработки и анализа научной информации) и их потребителей (исследовательские группы, научные организации, университеты, инновационные компании). К таким типовым услугам уже сейчас можно отнести подбор публикаций по заданной тематике, подготовку систематических обзоров и мета-анализов, выполнение патентных исследований и подготовку заявок на охрану интеллектуальной собственности.

Таким образом, современная научно-информационная инфраструктура не только улучшает качество научных результатов, но и способствует развитию рынка научных услуг, что, в свою очередь, ускоряет и удешевляет исследовательские проекты.

Контуры системы управления

Как уже отмечалось, основная сложность в управлении Программой состоит в необходимости консолидации интересов и актуализации направлений деятельности большого числа разнородных

участников информационного обмена. Любой научный журнал или информационная система является частью этой инфраструктуры, а любой исследователь (и не только исследователь) — ее пользователем. Выделим следующие *категории акторов* инфраструктуры:

1) поставщики ресурсов и услуг: научные издатели, университеты, научные организации;

2) потребители ресурсов и услуг: профессиональные сообщества исследователей, экспертные советы, университеты, научные организации, корпоративные R&D центры;

3) агрегаторы научной информации: РИНЦ, электронно-библиотечные системы, открытые архивы;

4) систематизаторы научной информации: институты научной информации, научные и научно-технические библиотеки;

5) операторы государственных информационных систем: Центр информационных технологий и систем органов исполнительной власти, Всероссийская аттестационная комиссия, Федеральный институт промышленной собственности, Российский научно-исследовательский институт экономики, политики и права в научно-технической сфере;

6) государственные регуляторы — федеральные органы исполнительной власти, чья деятельность существенно влияет на сферу научной-информации, например Минобрнауки России, Минцифры России, Минздрав России, Роспатент.

Поставщиками и потребителями научной информации только в России являются более 1 000 организаций. При этом значимые роли играют не только органы государственной власти и бюджетные учреждения, но и коммерческие, и некоммерческие организации. Для управления подобными сложными системами практикуется создание специального Консорциума, основными задачами которого являются (рис. 5):

1) разработка и актуализация Стратегии развития Национальной научно-информационной инфраструктуры;

2) планирование и контроль выполнения работ по реализации Стратегии;

3) координация деятельности *ключевых участников* Программы, к которым прежде всего относятся систематизаторы и агрегаторы научной информации, а также операторы государственных информационных систем;

4) содействие внедрению новых инфраструктурных и продуктовых ИТ-решений;

5) разработка и экспертиза нормативных документов, регулирующих деятельность в научно-информационной сфере, а также рекомендаций для увеличения ее эффективности;

6) распространение лучших методов и практик работы с научной информацией.

Для сбалансированной работы в состав Консорциума должны войти государственные регуляторы, Российская академия наук, ключевые участники, наиболее крупные поставщики научной информации и некоторые профессиональные научные сообщества. В деятельности Консорциума следует выделить предметные группы, в которых подготавливаются решения, специфичные для отдельных областей наук, а также 2 Совета.

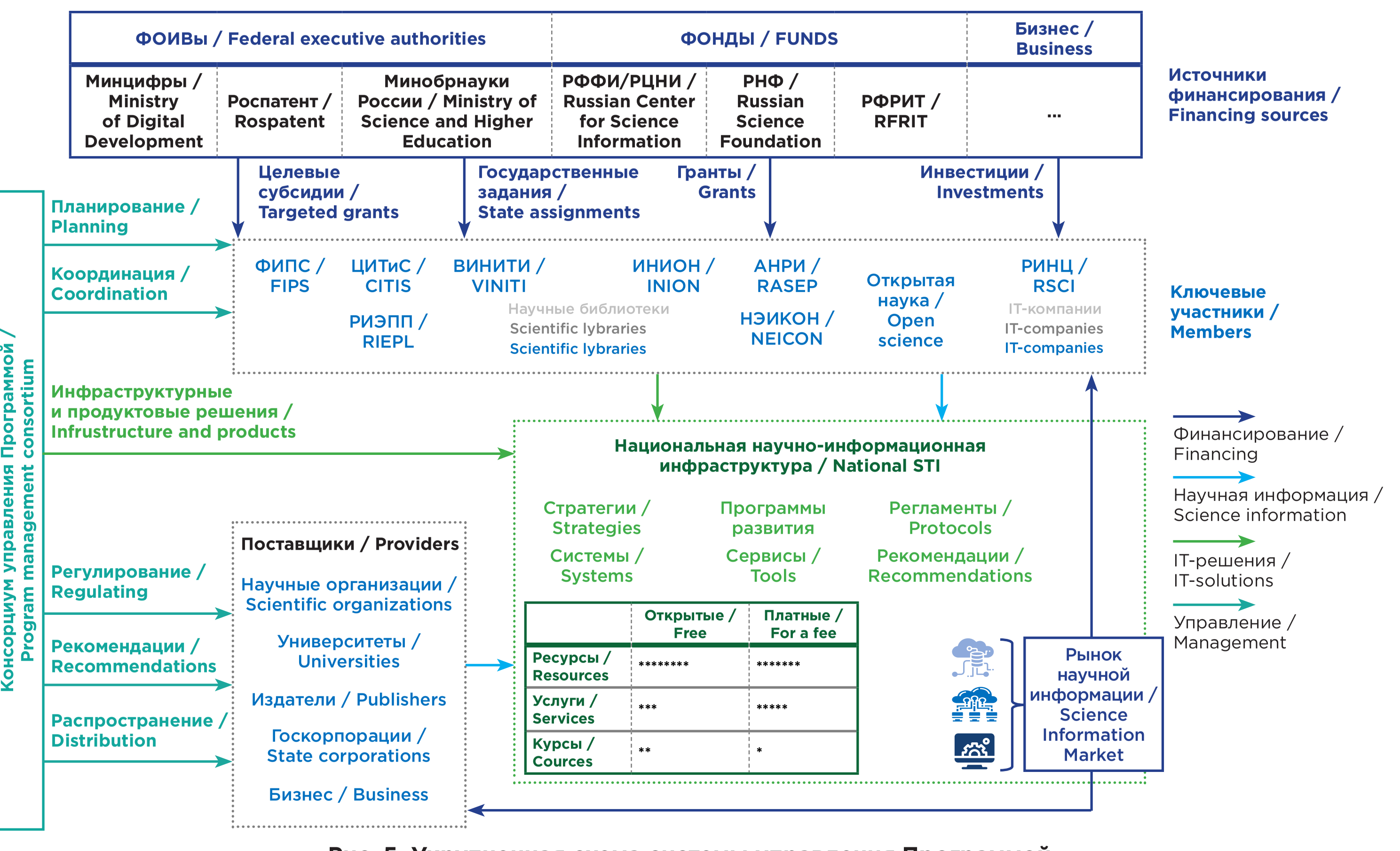


Рис. 5. Укрупненная схема системы управления Программой
Fig. 5. Schema of the Program management system

1. Координационный совет, отвечающий за координацию деятельности акторов, взаимодействие с органами государственной власти, разработку нормативных документов и принятие стратегических решений.

2. Технологический совет, отвечающий за разработку, развитие и поддержку инфраструктурных ИТ-решений (информационных ресурсов, систем и сервисов), разработку протоколов хранения и обмена научными данными, программных интерфейсов и рекомендаций по их использованию.

Финансовую модель управления Программой еще предстоит разработать, но уже сейчас можно предположить, что она не потребует значительных дополнительных средств из бюджета страны. Многие вопросы лежат в организационной плоскости, после решения которых уже имеющееся финансирование может быть использовано в целях Программы гораздо более эффективно.

Также могут быть целевым образом привлечены иные источники. Например, в сотрудничестве с Российским научным фондом может быть объявлен конкурс научных грантов на исследования, ориентированные на развитие методов искусственного интеллекта в задачах обработки научной информации. Российский фонд развития информационных технологий может предоставить целевые гранты на разработку новых инфраструктурных ИТ-решений. Российский фонд фундаментальных исследований в настоящее время преобразуется в Российский центр научной информации, после чего также займет свою функциональную нишу в этой инфраструктуре.

Одной из метрик эффективности программы станет производительность труда, что подразумевает не повышение, а снижение информационной нагрузки на работника. При этом мы понимаем, что любой переход на новые инструменты приводит к временному падению производительности труда в момент перехода — этот факт известен из опыта компаний, проводящих автоматизацию производств и коммерческих компаний. Эти метрики должны быть рассчитаны заранее, а программа должна реализовываться в тесной связи с фактическими показателями по этой и другим, не менее важным метрикам продуктивности.

Помня о традиционных страхах научного сообщества, необходимо отметить, что основная миссия Консорциума — эволюционный путь развития, а никак не «сломать и заново построить» (революция) или «зарегулировать» обмен научной информацией (бюрократизация). Деятельность Консорциума должна быть прежде всего направлена на увеличение количества и качества научной информации, расширение возможностей ее использования (включая машинную обработку), создание условий для развития рынка научной информации и разработку необходимых для этого ИТ-решений.

Заключение / Conclusion

Развитие Национальной научно-информационной инфраструктуры является очень сложной и масштабной задачей в качестве ответа на вызовы, связанные с новой парадигмой исследовательской

деятельности Data-Driven Science, с учетом исторически сложившихся в России проблем в сфере научных коммуникаций. Эта Программа должна основываться на Стратегии, разработка которой выполняется с учетом принципов коллегиальности, структурности, иерархии целеполагания и программного подхода. Ее целью должно быть развитие инфраструктуры рынка научной информации в соответствии с парадигмой Data-Driven Science, а основной миссией — создание «питательной среды» для формирования цепочек приращения ценности научной информации.

Результатом Программы должны стать новые высокоинтеллектуальные сервисы для информационной поддержки различных видов научной деятельности; развитие цифровых компетенций исследователей; значительное увеличение интенсивности использования научного контента; развитие рынка научно-информационных услуг; улучшения в системе разделения научного труда. Все это позволит университетам и научным организациям более интенсивно накапливать и использовать цифровой капитал [5], который уже не просто играет значимую роль в научно-технологическом развитии, а определяет конкурентоспособность субъектов этого процесса.

В заключение отметим, что при реализации такой масштабной Программы следует соблюдать принцип «поедания слона по кусочкам». Поскольку этот «слон» действительно большой, придется применять и комбинировать различные способы его «нарезки», такие как:

- по функциональным аспектам;
- по задачам и уровням ценности научной информации;
- по видам и источникам публикаций;
- по категориям акторов научно-информационной инфраструктуры;
- по преодолеваемым барьерам.

Однако нужно ли вообще начинать эту долгую «трапезу»? Или, может быть, лучше оставить все как есть? Отвечая на этот вопрос, можно провести аналогию между информационной инфраструктурой науки и транспортной инфраструктурой региона.

Если в регионе плохие дороги, неудобный общественный транспорт, неразвитое железнодорожное и авиасообщение, то жизнь от этого не остановится. Люди все равно будут перемещаться, грузы — доставляться, а бизнес — функционировать. Но это будет сопровождаться постоянными потерями времени и дополнительными расходами, что будет демотивировать пользователей инфраструктуры (никто не хочет лишний раз ехать по плохой дороге, а потом ремонтировать машину). Качество жизни в регионе будет оставаться низким, а туристическая и инвестиционная привлекательность — слабой.

Потери от плохой инфраструктуры могут быть очень большими, но при этом они часто оказываются недооцененными, поскольку эти потери очень трудно измерить из-за отсутствия базы для сравнения. Есть и множество примеров, когда вокруг новой инфраструктуры, отвечающей всем современным требованиям, выстраиваются целые города и экосистемы.

Аналогия с транспортной инфраструктурой здесь неслучайна. Научно-информационная инфраструктура в определенном смысле прокладывает «дороги» между «городами» исследователей, способствуя быстрой и качественной доставке основной ценности этого

«мира» — научного знания. В настоящее время наблюдается все больше признаков того, что сложившаяся инфраструктура «ветшает» и перестает соответствовать современным требованиям, которые к тому же резко возросли. На фоне других вызовов перед отечественной наукой эта проблема может казаться не столь значимой, однако без ее решения ответы на эти вызовы, скорее всего, окажутся неэффективными.

Для реализации Программы необходимо создание управляющего органа, который будет заниматься разработкой Стратегии, контролем соблюдения ее положений, согласованием требований держателей научно-исследовательской инфраструктуры, исследователей, предпринимателей; разработкой, а также координацией их деятельности. Это чрезвычайно трудная работа, требующая филигранной синхронизации требований, очень различающихся между собой по характеру деятельности субъектов. Сложность этой программы сопоставима с крупными национальными проектами вроде организации Олимпиады 2014 г. и трансформации энергетического рынка РФ РАО ЕЭС, но у нас просто нет другого выхода, если мы рассчитываем на то, что Россия удержит свои позиции в мире в качестве научной державы во время и после находящейся в разгаре научной и промышленной революции.

Список использованных источников

1. Electronic Journals and Changes in Scholarly Article Seeking and Reading Patterns / C. Tenopir [et al.] // Aslib Proceedings. 2009. Vol. 61, No. 1. P. 5—32. DOI: https://doi.org/10.1108/00012530910932267

2. Van Noorden R. Scientists May be Reaching a Peak in Reading Habits // Nature. 2014. DOI: https://doi.org/10.1038/nature.2014.14658

3. Антопольский А. Б. Информационная инфраструктура социально-гуманитарных наук в Евросоюзе // Наука и научная информация. 2021. Т. 4, № 1-2. С. 8—22. DOI: https://doi.org/10.24108/2658-3143-2021-4-1-2-18-32

4. Writer B. Lithium-Ion Batteries: A Machine-Generated Summary of Current Research. Heidelberg: Springer, 2019. 247 p. DOI: https://doi.org/10.1007/978-3-030-16800-1

5. Родионов И. И., Гиляревский Р. С., Цветкова В. В. Неоинформационная экономика и ее общество: тенденции развития. Москва, 2021. 391 с. DOI: https://doi.org/10.36535/978-5-600-02928-6

Информация об авторах

Гуськов Андрей Евгеньевич, кандидат технических наук, заведующий лабораторией наукометрии и научных коммуникаций, Российский научно-исследовательский институт экономики, политики и права в научно-технической сфере (127254, Россия, г. Москва, ул. Добролюбова, д. 20А), ORCID: https://orcid.org/0000-0002-1028-9958, guskov.andrey@gmail.com

Карауш Александр Сергеевич, кандидат технических наук, генеральный директор, Государственная публичная научно-техническая

библиотека России (123298, Россия, г. Москва, ул. 3-я Хорошевская, д. 17), ORCID: https://orcid.org/0000-0002-1589-8672, a@karaush.ru

Меньщиков Илья Евгеньевич, кандидат химических наук, заместитель руководителя инженерно-технического центра, Институт физической химии и электрохимии им. А. Н. Фрумкина Российской академии наук (119071, Россия, г. Москва, Ленинский пр-т., д. 31, к. 4), ORCID: https://orcid.org/0000-0001-8126-1023, i.menshchikov@gmail.com

Школин Андрей Вячеславович, кандидат химических наук, руководитель инженерно-технического центра, Институт физической химии и электрохимии им. А. Н. Фрумкина Российской академии наук, (119071, Россия, г. Москва, Ленинский пр-т., д. 31, к. 4), ORCID: https://orcid.org/0000-0001-6446-4419, shkolin@bk.ru

Недельский Виталий Олегович, кандидат политических наук, и. о. ректора, Горно-Алтайский государственный университет (649000, Россия, г. Горно-Алтайск, ул. Ленина, д. 1), ORCID: https://orcid.org/0000-0003-3346-8112, vn010470@gmail.com

Сабиров Денис Шамилевич, доктор химических наук, директор, Институт нефтехимии и катализа Уфимского федерального исследовательского центра РАН (450075, Россия, г. Уфа, пр-т Октября, д. 141), ORCID: https://orcid.org/0000-0002-6928-5070, sabirovdsh@mail.ru

Щукин Тимур Николаевич, кандидат психологических наук, модератор групповой работы, Московская школа управления «Сколково» (143025, Россия, Московская обл., ул. Новая, д. 100), ORCID: https://orcid.org/0000-0003-0783-7367, timurid@gmail.com

Заявленный вклад соавторов

Гуськов А. Е., Карауш А. С., Меньщиков И. Е., Школин А. В., Недельский В. О., Сабиров Д. Ш. — участие в постановке задачи, формулировании выводов и результатов исследования, критический анализ, подготовка текста; **Щукин Т. Н.** — участие в постановке задачи, формулировании выводов и результатов исследования, критический анализ, модерация обсуждений.

References

1. Tenopir C, King DW, Edwards S, Wu L. Electronic Journals and Changes in Scholarly Article Seeking and Reading Patterns. *Aslib Proceedings.* 2009;61(1):5-32. DOI: https://doi.org/10.1108/00012530910932267

2. Van Noorden R. Scientists May be Reaching a Peak in Reading Habits. *Nature.* 2014. DOI: https://doi.org/10.1038/nature.2014.14658

3. Antopolskiy AB. Scientific Infrastructure of Social Sciences and Humanities in European Union. *Scholarly Research and Information.* 2021;4(1-2):8-22. DOI: https://doi.org/10.24108/2658-3143-2021-4-1-2-18-32 (In Russ.)

4. Writer B. Lithium-ion Batteries: A Machine-Generated Summary of Current Research. Heidelberg: Springer, 2019. 247 p. DOI: https://doi.org/10.1007/978-3-030-16800-1

5. Rodionov II, Gilyarevskiy RS, Tsvetkova VV. Neo-Information Economics and Its Society: Development Trends. Moscow, 2021. 391 p. DOI: https://doi.org/10.36535/978-5-600-02928-6 (In Russ.)

Information about the authors

Andrey E. Guskov, Ph.D. (Technical Sciences), Head of the Laboratory of Scientometrics and Scientific Communications, Russian Research Institute of Economics, Politics and Law in Science and Technology (20A Dobrolyubova St., Moscow 127254, Russia), ORCID: https://orcid.org/0000-0002-1028-9958, guskov.andrey@gmail.com

Aleksandr S. Karaush, Ph.D. (Technical Sciences), Executive Director, Russian National Public Library for Science and Technology (17 3rd Khoroshevskaya St., Moscow 123298, Russia), ORCID: https://orcid.org/0000-0002-1589-8672, a@karaush.ru

Ilya E. Menshchikov, Ph.D. (Chemical Sciences), Deputy Head of the Engineering Center of the IPCE RAS, A. N. Frumkin Institute of Physical Chemistry and Electrochemistry RAS (31/4 Leninsky Prospect, Moscow 119071, Russia), ORCID: https://orcid.org/0000-0001-8126-1023, i.menshchikov@gmail.com

Andrei V. Shkolin, Ph.D. (Chemical Sciences), Head of the Engineering Center of the IPCE RAS, A. N. Frumkin Institute of Physical Chemistry and Electrochemistry RAS (31/4 Leninsky Prospect, Moscow 119071, Russia), ORCID: https://orcid.org/0000-0001-6446-4419, shkolin@bk.ru

Vitaly O. Nedelskiy, Ph.D. (Political Science), Acting Rector, Gorno-Altai State University (1 Lenina St., Gorno-Altaisk 649000, Russia), ORCID: https://orcid.org/0000-0003-3346-8112, vn010470@gmail.com

Denis S. Sabirov, Dr.Sci. (Chemical Sciences), Head, Institute of Petrochemistry and Catalysis, Ufa Federal Research Centre of RAS (41 Oktyabrya Prospect, Ufa 450075, Russia), ORCID: https://orcid.org/0000-0002-6928-5070, sabirovdsh@mail.ru

Timur N. Shchukin, Ph.D. (Psychologocal Sciences), Moderator, Moscow School of management "Skolkovo" (100 Novaya St., Moscow Region 143026, Russia), ORCID: https://orcid.org/0000-0003-0783-7367, timurid@gmail.com

Contribution of the authors

A. E. Guskov, A. S. Karaush, I. E. Menshchikov, A. V. Shkolin, V. O. Nedelskiy, D. S. Sabirov — participation in problem statement, formulation of conclusions and research results, critical analysis, preparing the text; **T. N. Shchukin** — participation in problem statement, formulation of conclusions and research results, critical analysis, moderation of discussions.

Авторы заявляют об отсутствии конфликта интересов.
The authors declare no conflict of interests.

Поступила 27.07.2022 *Submitted 27.07.2022*
Одобрена 05.09.2022 *Approved 05.09.2022*
Принята 07.09.2022 *Accepted 07.09.2022*